



Model Fine-Tuning, Reinforcement, and Synthetic Data Governance Standard

Defines data, supervised tuning, PEFT, preference optimization, reinforcement, synthetic-data generation, evaluation, provenance, safety, release, and rollback.

MYTHOS-AI-0009

Candidate Mythos Standard / 2.0.0-candidate

Published 2026-07-26 | Review 2026-10-26 | Public

Mythos Systems

Permanent Identity and Edition Record

Permanent document ID	MYTHOS-AI-0009
Canonical title	Model Fine-Tuning, Reinforcement, and Synthetic Data Governance Standard
Publication class	Standards & Benchmarks
Primary domain	MYTHOS-AI - Artificial Intelligence & Agentic Systems
Version	2.0.0-candidate
Status	Candidate Mythos Standard
Publication date	2026-07-26
Scheduled review	2026-10-26
Publisher	Mythos Systems
Owner	Aaron Stovall
Classification	Public
Prior edition	1.0.0-candidate / 2026-07-24
Supersession	This edition supersedes the prior candidate while preserving the permanent identity and historical source.

Identity doctrine

The permanent document ID is immutable. Production sequence labels are not part of the public identity. Atomic standards are authoritative; portfolios, workbooks, and machine-readable exports are generated views.

Lineage clarification

Corrects the earlier public title token 'RLEF' to 'Reinforcement' and expands the normative scope without changing the permanent document identity.

Normative Language and Applicability

The key words MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL indicate the strength of provisions in this Mythos standard. The publication is a Mythos-authored candidate standard and does not claim approval by the cited external authorities.

Applicability rule

Every control is applicable unless the organization documents a specific architectural or lifecycle reason that the subject is absent. N/A decisions MUST name the decision owner, evidence, affected scope, review date, and triggers that invalidate the exclusion.

Conformance evidence hierarchy

Direct deterministic proof: schemas, tests, signatures, state reconciliation, access decisions, build or runtime evidence.

Reproducible technical evidence: benchmark fixtures, traces, artifacts, profiles, logs, metrics, and environment manifests.

Independent assessment: qualified human or separate evaluator using an explicit rubric and disclosed uncertainty.

Attestation or supplier claim: informative unless independently verified and current for the deployed configuration.

Demonstration or narrative: insufficient by itself for a conforming score above 2.

Critical-control doctrine

Critical controls cannot be averaged away. A failed or stale critical control constrains the maximum conformance level regardless of the aggregate score.

Scope, Audience, and Engineering Principles

Purpose. Defines data, supervised tuning, PEFT, preference optimization, reinforcement, synthetic-data generation, evaluation, provenance, safety, release, and rollback.

In-scope systems. full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation.

Primary audience. model engineers, data teams, alignment teams, evaluators, security teams, platform teams, and governance leaders.

Required engineering principles

Model capability and complete system behavior **MUST** be measured separately.

Human, legal, production, financial, identity, and governance authority **MUST** remain external to model confidence or conversational fluency.

Evaluation **MUST** include expected, prohibited, adversarial, degraded, recovery, and lifecycle-change conditions.

Source authority, data rights, permissions, versions, and freshness **MUST** be visible in evidence.

Critical outcomes **MUST** be supported by deterministic validators or independent review wherever feasible.

The organization **MUST** preserve rollback, revocation, correction, deletion, and retirement paths.

Optimization for score, latency, throughput, tokens, or cost **MUST NOT** hide safety, quality, privacy, accessibility, or reliability regressions.

Exclusions

This standard does not certify a model, provider, framework, benchmark, dataset, or product. Conformance is limited to the declared system, use case, version, environment, evidence period, and assessment scope.

Conformance and Scoring Model

Level	Weighted threshold	Critical-control condition	Meaning
No conformance	< 50	Any state	Materially incomplete, unverified, or unsafe.
Foundation	>= 50	No critical control scored 0	Defined baseline with partial implementation and evidence.
Operational	>= 65	All applicable critical controls >= 3	Implemented and operationally tested.
Assured	>= 80	All applicable critical controls >= 4	Independently verified with adversarial and recovery evidence.
Continuously Assured	>= 90	All applicable critical controls = 5	Continuous regression, monitoring, freshness, and incident learning.

Weighted score

For applicable controls: weighted contribution = (control score / 5) x control weight. N/A controls are removed from the denominator only after an approved applicability decision. The final score is normalized to 100.

A conformance claim MUST include scope, system and model versions, assessment date, evidence window, evaluator identity, scores by family, critical-control status, unresolved findings, exceptions, limitations, and next review.

Control Score Interpretation

Score	Maturity	Required evidence state
0	Not implemented	Not implemented: the assessed control is absent, materially unknown, or contradicted by evidence.
1	Initial	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Nonconformity classes

Critical: credible risk of serious harm, unauthorized high-impact action, material security or privacy failure, false assurance, or inability to recover.

Major: requirement absent or ineffective across a meaningful part of scope, with no sufficient compensating control.

Minor: localized or low-impact deficiency that does not invalidate the overall control objective.

Observation: improvement opportunity or emerging risk without current nonconformity.

Score validity

A score expires when evidence exceeds its approved freshness period or a material change invalidates the assessment. Current operation is not inferred from a historic assessment.

Standard Architecture and Control Model

01	Govern	Scope, ownership, authority, policy, impacts, risk tolerance, exceptions, and lifecycle decisions.
02	Map	System boundaries, actors, models, data, prompts, tools, interfaces, populations, dependencies, and failure domains.
03	Measure	Representative and hidden evaluation, deterministic validation, human or model judging, uncertainty, performance, and cost.
04	Manage	Release gates, least privilege, monitoring, incident response, correction, rollback, revocation, deletion, and retirement.
05	Prove	Traceable evidence bundles connecting claims to configurations, tests, artifacts, effects, decisions, and user outcomes.

Assessment unit

The assessment unit is the declared AI system or workflow, not the model name alone. It includes prompts, retrieval, memory, tools, policies, interface, evaluators, infrastructure, human oversight, and operating environment.

Benchmark Dimensions and Weights

Dimension	Weight	Assessment emphasis
Objective and impact	12%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for objective and impact.
Data authority and quality	23%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for data authority and quality.
Adaptation configuration	17%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for adaptation configuration.
Preferences and reinforcement	15%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for preferences and reinforcement.
Training systems and provenance	12%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for training systems and provenance.
Evaluation and safety	15%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for evaluation and safety.
Release and lifecycle	6%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for release and lifecycle.

Benchmark scores **MUST** be reported by dimension and by critical-control status. A single aggregate ranking **MUST NOT** be used to conceal weak safety, authority, privacy, recovery, or evidence performance.

Contents and Control Index

[Permanent identity and edition record](#)

[Normative language and applicability](#)

[Scope, audience, and engineering principles](#)

[Conformance and scoring model](#)

[Standard architecture and control model](#)

[Benchmark dimensions and weights](#)

[Control summary matrix](#)

MYTHOS-AI-0009-C01 - Adaptation Objective and Method Decision

MYTHOS-AI-0009-C02 - Impact, Risk, and Approval Classification

MYTHOS-AI-0009-C03 - Training Data Rights and Authority

MYTHOS-AI-0009-C04 - Dataset Manifest, Lineage, and Versioning

MYTHOS-AI-0009-C05 - Privacy, Secrets, and Sensitive Data Controls

MYTHOS-AI-0009-C06 - Data Quality, Balance, and Deduplication

MYTHOS-AI-0009-C07 - Synthetic Data Generation Contract

MYTHOS-AI-0009-C08 - Curriculum, Sampling, and Difficulty Governance

MYTHOS-AI-0009-C09 - Supervised Fine-Tuning Configuration

MYTHOS-AI-0009-C10 - PEFT, LoRA, and Quantized Adaptation

MYTHOS-AI-0009-C11 - Preference Data and Annotation Governance

MYTHOS-AI-0009-C12 - Reward Model and Reinforcement Objective Governance

MYTHOS-AI-0009-C13 - Direct Preference and Alternative Optimization

MYTHOS-AI-0009-C14 - Distributed Training and Resource Control

MYTHOS-AI-0009-C15 - Checkpoint, Resume, and Artifact Provenance

MYTHOS-AI-0009-C16 - Evaluation Separation and Base-Tuned Comparison

MYTHOS-AI-0009-C17 - Forgetting, Overfitting, and Distribution Shift

MYTHOS-AI-0009-C18 - Safety, Security, and Memorization Evaluation

MYTHOS-AI-0009-C19 - Release, Monitoring, Rollback, and Retirement

Control Summary Matrix

Control	Requirement	Family	Wt.	Crit.
C01	Adaptation Objective and Method Decision	Objective and impact	6	YES
C02	Impact, Risk, and Approval Classification	Objective and impact	6	YES
C03	Training Data Rights and Authority	Data authority and quality	5	YES
C04	Dataset Manifest, Lineage, and Versioning	Data authority and quality	5	YES
C05	Privacy, Secrets, and Sensitive Data Controls	Data authority and quality	5	YES
C06	Data Quality, Balance, and Deduplication	Data authority and quality	4	YES
C07	Synthetic Data Generation Contract	Data authority and quality	4	YES
C08	Curriculum, Sampling, and Difficulty Governance	Adaptation configuration	6	-
C09	Supervised Fine-Tuning Configuration	Adaptation configuration	6	YES
C10	PEFT, LoRA, and Quantized Adaptation	Adaptation configuration	5	-

Control Summary Matrix - Continued

Control	Requirement	Family	Wt.	Crit.
C11	Preference Data and Annotation Governance	Preferences and reinforcement	5	YES
C12	Reward Model and Reinforcement Objective Governance	Preferences and reinforcement	5	YES
C13	Direct Preference and Alternative Optimization	Preferences and reinforcement	5	-
C14	Distributed Training and Resource Control	Training systems and provenance	6	YES
C15	Checkpoint, Resume, and Artifact Provenance	Training systems and provenance	6	YES
C16	Evaluation Separation and Base-Tuned Comparison	Evaluation and safety	5	YES
C17	Forgetting, Overfitting, and Distribution Shift	Evaluation and safety	5	YES
C18	Safety, Security, and Memorization Evaluation	Evaluation and safety	5	YES
C19	Release, Monitoring, Rollback, and Retirement	Release and lifecycle	6	YES

Weight integrity

The 19 applicable control weights sum to 100. Family weights match the benchmark dimension model. Critical controls are highlighted in the atomic control pages and assessment workbook.

MYTHOS-AI-0009-C01

Adaptation Objective and Method Decision

Family: Objective and impact | Weight: 6 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain adaptation objective and method decision for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. The operating contract must define intended users, decisions, prohibited uses, affected systems, consequences, risk tolerance, terminal conditions, and accountable authority. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that adaptation objective and method decision is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for adaptation objective and method decision covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

The operating contract must define intended users, decisions, prohibited uses, affected systems, consequences, risk tolerance, terminal conditions, and accountable authority.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C01

Adaptation Objective and Method Decision

Family: Objective and impact | Weight: 6 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for adaptation objective and method decision.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

negative, boundary, degraded, and recovery test results

operational telemetry and current-state verification

Required metrics

applicable control coverage

critical-failure count

evidence freshness

open nonconformities

Score anchors

0	Not implemented: Adaptation Objective and Method Decision is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for adaptation objective and method decision.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C02

Impact, Risk, and Approval Classification

Family: Objective and impact | Weight: 6 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain impact, risk, and approval classification for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Authority must be enforceable outside the model through stable identities, least privilege, scoped credentials, separation of duties, revocation, and auditable approval. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that impact, risk, and approval classification is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for impact, risk, and approval classification covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Authority must be enforceable outside the model through stable identities, least privilege, scoped credentials, separation of duties, revocation, and auditable approval.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C02

Impact, Risk, and Approval Classification

Family: Objective and impact | Weight: 6 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for impact, risk, and approval classification.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Impact, Risk, and Approval Classification is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for impact, risk, and approval classification.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C03

Training Data Rights and Authority

Family: Data authority and quality | Weight: 5 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain training data rights and authority for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Authority must be enforceable outside the model through stable identities, least privilege, scoped credentials, separation of duties, revocation, and auditable approval. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that training data rights and authority is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for training data rights and authority covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Authority must be enforceable outside the model through stable identities, least privilege, scoped credentials, separation of duties, revocation, and auditable approval.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C03

Training Data Rights and Authority

Family: Data authority and quality | Weight: 5 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for training data rights and authority.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

dataset or source manifest with rights, lineage, quality, and split records

privacy, retention, deletion, and contamination review

Required metrics

lineage completeness

rights-review coverage

duplicate or contamination rate

quality-slice pass rate

deletion-verification rate

Score anchors

0	Not implemented: Training Data Rights and Authority is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for training data rights and authority.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C04

Dataset Manifest, Lineage, and Versioning

Family: Data authority and quality | Weight: 5 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain dataset manifest, lineage, and versioning for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that dataset manifest, lineage, and versioning is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for dataset manifest, lineage, and versioning covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C04

Dataset Manifest, Lineage, and Versioning

Family: Data authority and quality | Weight: 5 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for dataset manifest, lineage, and versioning.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

dataset or source manifest with rights, lineage, quality, and split records

privacy, retention, deletion, and contamination review

Required metrics

lineage completeness

rights-review coverage

duplicate or contamination rate

quality-slice pass rate

deletion-verification rate

Score anchors

0	Not implemented: Dataset Manifest, Lineage, and Versioning is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for dataset manifest, lineage, and versioning.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C05

Privacy, Secrets, and Sensitive Data Controls

Family: Data authority and quality | Weight: 5 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain privacy, secrets, and sensitive data controls for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that privacy, secrets, and sensitive data controls is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for privacy, secrets, and sensitive data controls covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C05

Privacy, Secrets, and Sensitive Data Controls

Family: Data authority and quality | Weight: 5 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for privacy, secrets, and sensitive data controls.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

dataset or source manifest with rights, lineage, quality, and split records

privacy, retention, deletion, and contamination review

Required metrics

lineage completeness

rights-review coverage

duplicate or contamination rate

quality-slice pass rate

deletion-verification rate

Score anchors

0	Not implemented: Privacy, Secrets, and Sensitive Data Controls is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for privacy, secrets, and sensitive data controls.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C06

Data Quality, Balance, and Deduplication

Family: Data authority and quality | Weight: 4 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain data quality, balance, and deduplication for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that data quality, balance, and deduplication is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for data quality, balance, and deduplication covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C06

Data Quality, Balance, and Deduplication

Family: Data authority and quality | Weight: 4 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for data quality, balance, and deduplication.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

dataset or source manifest with rights, lineage, quality, and split records

privacy, retention, deletion, and contamination review

Required metrics

lineage completeness

rights-review coverage

duplicate or contamination rate

quality-slice pass rate

deletion-verification rate

Score anchors

0	Not implemented: Data Quality, Balance, and Deduplication is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for data quality, balance, and deduplication.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C07

Synthetic Data Generation Contract

Family: Data authority and quality | Weight: 4 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain synthetic data generation contract for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that synthetic data generation contract is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for synthetic data generation contract covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C07

Synthetic Data Generation Contract

Family: Data authority and quality | Weight: 4 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for synthetic data generation contract.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

dataset or source manifest with rights, lineage, quality, and split records

privacy, retention, deletion, and contamination review

Required metrics

lineage completeness

rights-review coverage

duplicate or contamination rate

quality-slice pass rate

deletion-verification rate

Score anchors

0	Not implemented: Synthetic Data Generation Contract is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for synthetic data generation contract.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C08

Curriculum, Sampling, and Difficulty Governance

Family: Adaptation configuration | Weight: 6 | Critical: NO

Normative requirement

The organization **MUST** establish, implement, verify, and maintain curriculum, sampling, and difficulty governance for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. The control must identify accountable ownership, authoritative inputs, expected outputs, operating limits, dependencies, failure behavior, evidence, and reassessment triggers. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that curriculum, sampling, and difficulty governance is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for curriculum, sampling, and difficulty governance covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

The control must identify accountable ownership, authoritative inputs, expected outputs, operating limits, dependencies, failure behavior, evidence, and reassessment triggers.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C08

Curriculum, Sampling, and Difficulty Governance

Family: Adaptation configuration | Weight: 6 | Critical: NO

Assessment procedure

- Examine the approved design, applicability decision, responsible roles, and current implementation for curriculum, sampling, and difficulty governance.
- Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.
- Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.
- Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.
- Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.
- Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Curriculum, Sampling, and Difficulty Governance is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

- The organization cannot identify an authoritative owner, current scope, or complete implementation for curriculum, sampling, and difficulty governance.
- Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.
- Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.
- Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C09

Supervised Fine-Tuning Configuration

Family: Adaptation configuration | Weight: 6 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain supervised fine-tuning configuration for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. The inventory must identify versions, owners, suppliers, deployment locations, interfaces, dependencies, trust boundaries, and supported lifecycle states. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that supervised fine-tuning configuration is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for supervised fine-tuning configuration covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

The inventory must identify versions, owners, suppliers, deployment locations, interfaces, dependencies, trust boundaries, and supported lifecycle states.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C09

Supervised Fine-Tuning Configuration

Family: Adaptation configuration | Weight: 6 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for supervised fine-tuning configuration.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

negative, boundary, degraded, and recovery test results

operational telemetry and current-state verification

Required metrics

applicable control coverage

critical-failure count

evidence freshness

open nonconformities

Score anchors

0	Not implemented: Supervised Fine-Tuning Configuration is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for supervised fine-tuning configuration.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C10

PEFT, LoRA, and Quantized Adaptation

Family: Adaptation configuration | Weight: 5 | Critical: NO

Normative requirement

The organization **MUST** establish, implement, verify, and maintain peft, lora, and quantized adaptation for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. The control must identify accountable ownership, authoritative inputs, expected outputs, operating limits, dependencies, failure behavior, evidence, and reassessment triggers. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that peft, lora, and quantized adaptation is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for peft, lora, and quantized adaptation covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

The control must identify accountable ownership, authoritative inputs, expected outputs, operating limits, dependencies, failure behavior, evidence, and reassessment triggers.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C10

PEFT, LoRA, and Quantized Adaptation

Family: Adaptation configuration | Weight: 5 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for peft, lora, and quantized adaptation.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

negative, boundary, degraded, and recovery test results

operational telemetry and current-state verification

Required metrics

applicable control coverage

critical-failure count

evidence freshness

open nonconformities

Score anchors

0	Not implemented: PEFT, LoRA, and Quantized Adaptation is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for peft, lora, and quantized adaptation.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C11

Preference Data and Annotation Governance

Family: Preferences and reinforcement | Weight: 5 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain preference data and annotation governance for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that preference data and annotation governance is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for preference data and annotation governance covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C11

Preference Data and Annotation Governance

Family: Preferences and reinforcement | Weight: 5 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for preference data and annotation governance.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner
versioned configuration, schema, policy, model, prompt, dataset, or architecture record
reproducible test plan with expected and observed results
traceable decision, exception, approval, and remediation records
dataset or source manifest with rights, lineage, quality, and split records
privacy, retention, deletion, and contamination review

Required metrics

lineage completeness
rights-review coverage
duplicate or contamination rate
quality-slice pass rate
deletion-verification rate

Score anchors

0	Not implemented: Preference Data and Annotation Governance is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for preference data and annotation governance.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C12

Reward Model and Reinforcement Objective Governance

Family: Preferences and reinforcement | Weight: 5 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain reward model and reinforcement objective governance for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. The operating contract must define intended users, decisions, prohibited uses, affected systems, consequences, risk tolerance, terminal conditions, and accountable authority. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that reward model and reinforcement objective governance is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for reward model and reinforcement objective governance covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

The operating contract must define intended users, decisions, prohibited uses, affected systems, consequences, risk tolerance, terminal conditions, and accountable authority.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C12

Reward Model and Reinforcement Objective Governance

Family: Preferences and reinforcement | Weight: 5 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for reward model and reinforcement objective governance.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

negative, boundary, degraded, and recovery test results

operational telemetry and current-state verification

Required metrics

applicable control coverage

critical-failure count

evidence freshness

open nonconformities

Score anchors

0	Not implemented: Reward Model and Reinforcement Objective Governance is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for reward model and reinforcement objective governance.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C13

Direct Preference and Alternative Optimization

Family: Preferences and reinforcement | Weight: 5 | Critical: NO

Normative requirement

The organization **MUST** establish, implement, verify, and maintain direct preference and alternative optimization for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Training governance must bind objective, data, tokenizer, base model, method, hyperparameters, checkpoints, evaluation, safety, provenance, approval, deployment, and rollback. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that direct preference and alternative optimization is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for direct preference and alternative optimization covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Training governance must bind objective, data, tokenizer, base model, method, hyperparameters, checkpoints, evaluation, safety, provenance, approval, deployment, and rollback.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C13

Direct Preference and Alternative Optimization

Family: Preferences and reinforcement | Weight: 5 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for direct preference and alternative optimization.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner
versioned configuration, schema, policy, model, prompt, dataset, or architecture record
reproducible test plan with expected and observed results
traceable decision, exception, approval, and remediation records
negative, boundary, degraded, and recovery test results
operational telemetry and current-state verification

Required metrics

applicable control coverage
critical-failure count
evidence freshness
open nonconformities

Score anchors

0	Not implemented: Direct Preference and Alternative Optimization is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for direct preference and alternative optimization.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C14

Distributed Training and Resource Control

Family: Training systems and provenance | Weight: 6 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain distributed training and resource control for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that distributed training and resource control is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for distributed training and resource control covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C14

Distributed Training and Resource Control

Family: Training systems and provenance | Weight: 6 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for distributed training and resource control.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

dataset or source manifest with rights, lineage, quality, and split records

privacy, retention, deletion, and contamination review

Required metrics

lineage completeness

rights-review coverage

duplicate or contamination rate

quality-slice pass rate

deletion-verification rate

Score anchors

0	Not implemented: Distributed Training and Resource Control is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for distributed training and resource control.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C15

Checkpoint, Resume, and Artifact Provenance

Family: Training systems and provenance | Weight: 6 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain checkpoint, resume, and artifact provenance for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Evidence must correlate system identity, version, configuration, inputs, outputs, actions, approvals, metrics, artifacts, findings, exceptions, and final outcomes with integrity protection. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that checkpoint, resume, and artifact provenance is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for checkpoint, resume, and artifact provenance covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Evidence must correlate system identity, version, configuration, inputs, outputs, actions, approvals, metrics, artifacts, findings, exceptions, and final outcomes with integrity protection.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C15

Checkpoint, Resume, and Artifact Provenance

Family: Training systems and provenance | Weight: 6 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for checkpoint, resume, and artifact provenance.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner

versioned configuration, schema, policy, model, prompt, dataset, or architecture record

reproducible test plan with expected and observed results

traceable decision, exception, approval, and remediation records

negative, boundary, degraded, and recovery test results

operational telemetry and current-state verification

Required metrics

applicable control coverage

critical-failure count

evidence freshness

open nonconformities

Score anchors

0	Not implemented: Checkpoint, Resume, and Artifact Provenance is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for checkpoint, resume, and artifact provenance.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C16

Evaluation Separation and Base-Tuned Comparison

Family: Evaluation and safety | Weight: 5 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain evaluation separation and base-tuned comparison for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Authority must be enforceable outside the model through stable identities, least privilege, scoped credentials, separation of duties, revocation, and auditable approval. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that evaluation separation and base-tuned comparison is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for evaluation separation and base-tuned comparison covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Authority must be enforceable outside the model through stable identities, least privilege, scoped credentials, separation of duties, revocation, and auditable approval.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C16

Evaluation Separation and Base-Tuned Comparison

Family: Evaluation and safety | Weight: 5 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for evaluation separation and base-tuned comparison.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner
versioned configuration, schema, policy, model, prompt, dataset, or architecture record
reproducible test plan with expected and observed results
traceable decision, exception, approval, and remediation records
negative, boundary, degraded, and recovery test results
operational telemetry and current-state verification

Required metrics

task success with confidence interval
critical-failure rate
slice variance
reproducibility delta
human-review burden

Score anchors

0	Not implemented: Evaluation Separation and Base-Tuned Comparison is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for evaluation separation and base-tuned comparison.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C17

Forgetting, Overfitting, and Distribution Shift

Family: Evaluation and safety | Weight: 5 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain forgetting, overfitting, and distribution shift for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. The control must identify accountable ownership, authoritative inputs, expected outputs, operating limits, dependencies, failure behavior, evidence, and reassessment triggers. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that forgetting, overfitting, and distribution shift is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for forgetting, overfitting, and distribution shift covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

The control must identify accountable ownership, authoritative inputs, expected outputs, operating limits, dependencies, failure behavior, evidence, and reassessment triggers.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C17

Forgetting, Overfitting, and Distribution Shift

Family: Evaluation and safety | Weight: 5 | Critical: YES

Assessment procedure

- Examine the approved design, applicability decision, responsible roles, and current implementation for forgetting, overfitting, and distribution shift.
- Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.
- Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.
- Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.
- Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.
- Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Forgetting, Overfitting, and Distribution Shift is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

- The organization cannot identify an authoritative owner, current scope, or complete implementation for forgetting, overfitting, and distribution shift.
- Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.
- Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.
- Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C18

Safety, Security, and Memorization Evaluation

Family: Evaluation and safety | Weight: 5 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain safety, security, and memorization evaluation for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Evaluation must use representative and hidden cases, reproducible configuration, deterministic validators where possible, statistical uncertainty, critical-failure gates, and slice-level reporting. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that safety, security, and memorization evaluation is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for safety, security, and memorization evaluation covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Evaluation must use representative and hidden cases, reproducible configuration, deterministic validators where possible, statistical uncertainty, critical-failure gates, and slice-level reporting.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C18

Safety, Security, and Memorization Evaluation

Family: Evaluation and safety | Weight: 5 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for safety, security, and memorization evaluation.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

approved control design or operating contract with accountable owner
versioned configuration, schema, policy, model, prompt, dataset, or architecture record
reproducible test plan with expected and observed results
traceable decision, exception, approval, and remediation records
adversarial case corpus with campaign provenance
critical-failure findings, control response, and retest evidence

Required metrics

task success with confidence interval
critical-failure rate
slice variance
reproducibility delta
human-review burden

Score anchors

0	Not implemented: Safety, Security, and Memorization Evaluation is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for safety, security, and memorization evaluation.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0009-C19

Release, Monitoring, Rollback, and Retirement

Family: Release and lifecycle | Weight: 6 | Critical: YES

Normative requirement

The organization **MUST** establish, implement, verify, and maintain release, monitoring, rollback, and retirement for all in-scope full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation. Recovery must cover safe stop, checkpoint integrity, rollback or compensation, credential revocation, artifact restoration, evidence preservation, requalification, and controlled retirement. The control **MUST** define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization **SHOULD** enforce the requirement outside the model or agent. Any residual manual dependency **MUST** be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that release, monitoring, rollback, and retirement is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

Establish a versioned control contract for release, monitoring, rollback, and retirement covering full fine-tuning, supervised fine-tuning, adapters, LoRA, QLoRA, preference optimization, reward models, reinforcement learning, synthetic data, and continual adaptation and name the accountable owner, approver, assessor, and affected stakeholders.

Recovery must cover safe stop, checkpoint integrity, rollback or compensation, credential revocation, artifact restoration, evidence preservation, requalification, and controlled retirement.

Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.

Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.

Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.

Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0009-C19

Release, Monitoring, Rollback, and Retirement

Family: Release and lifecycle | Weight: 6 | Critical: YES

Assessment procedure

- Examine the approved design, applicability decision, responsible roles, and current implementation for release, monitoring, rollback, and retirement.
- Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.
- Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.
- Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.
- Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.
- Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- tool-call and external-effect receipts
- failure-injection, idempotency, rollback, and post-change validation results

Required metrics

- recovery success rate
- duplicate-effect rate
- time to safe state
- residual-effect count
- evidence preservation rate

Score anchors

0	Not implemented: Release, Monitoring, Rollback, and Retirement is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

- The organization cannot identify an authoritative owner, current scope, or complete implementation for release, monitoring, rollback, and retirement.
- Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.
- Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.
- Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

Required Benchmark Scenarios

MYTHOS-AI-0009-B01 - Representative production task

Demonstrate the declared use case using current configuration and representative inputs. Apply the scenario specifically to model adaptation governance and preserve raw evidence.

MYTHOS-AI-0009-B02 - Hidden critical holdout

Execute high-impact cases withheld from development, tuning, and prompt iteration. Apply the scenario specifically to model adaptation governance and preserve raw evidence.

MYTHOS-AI-0009-B03 - Malformed and adversarial inputs

Test schema violations, ambiguity, direct or indirect manipulation, and unsafe instructions. Apply the scenario specifically to model adaptation governance and preserve raw evidence.

MYTHOS-AI-0009-B04 - Long-horizon or multi-step execution

Measure compounding error, context loss, looping, premature completion, and recovery. Apply the scenario specifically to model adaptation governance and preserve raw evidence.

MYTHOS-AI-0009-B05 - Dependency and tool failure

Inject model, network, data, tool, credential, evaluator, or service failure. Apply the scenario specifically to model adaptation governance and preserve raw evidence.

Required Benchmark Scenarios - Continued

MYTHOS-AI-0009-B06 - Resource pressure and overload

Exercise concurrency, long inputs, queueing, rate limits, memory, tokens, and cost limits. Apply the scenario specifically to model adaptation governance and preserve raw evidence.

MYTHOS-AI-0009-B07 - Privacy and cross-boundary test

Attempt cross-tenant, secret, personal-data, restricted-source, cache, and egress violations. Apply the scenario specifically to model adaptation governance and preserve raw evidence.

MYTHOS-AI-0009-B08 - Rollback, revocation, or restore

Prove safe stop, compensation, revocation, prior-state restoration, and residual-effect reconciliation. Apply the scenario specifically to model adaptation governance and preserve raw evidence.

MYTHOS-AI-0009-B09 - Human intervention and disagreement

Test approval, interruption, correction, appeal, assessor disagreement, and minority findings. Apply the scenario specifically to model adaptation governance and preserve raw evidence.

MYTHOS-AI-0009-B10 - Material change and regression

Change model, prompt, dataset, tool, provider, runtime, or policy and run requalification gates. Apply the scenario specifically to model adaptation governance and preserve raw evidence.

Conformance Report and Evidence Bundle

Permanent document ID, standard version, assessment version, system and use-case scope.

System, model, provider, prompt, data, retrieval, memory, tool, evaluator, runtime, repository, and infrastructure identities.

Applicability decisions and normalized denominator.

Control-level scores, weights, critical status, evidence links, assessor, date, confidence, and finding disposition.

Benchmark scenario results with raw artifacts, deterministic validators, judge or expert records, uncertainty, failures, and limitations.

Family and total score, achieved conformance level, unresolved critical or major findings, approved exceptions, expiry, and next review.

Release, rollback, revocation, deletion, and retirement evidence where applicable.

Evidence bundle integrity

The evidence bundle SHOULD use content hashes, signatures or protected repository history, stable artifact IDs, access controls, retention policy, and independent verification. Evidence links alone are insufficient when the referenced object can change without detection.

Exceptions, Waivers, and Compensating Controls

An exception **MUST NOT** be used to relabel a failed requirement as conforming. Exceptions document accepted residual risk for a bounded scope and time; the underlying control score remains evidence-based.

Identify the exact control, system scope, reason, affected users or assets, and current score.

Describe the missing requirement, residual risk, foreseeable failure, and affected decisions.

Define compensating controls, validation evidence, monitoring, owner, approver, start date, expiry, and remediation plan.

Obtain approval from an authority independent of the implementation owner for critical or high-impact exceptions.

Reassess on incident, material change, evidence expiry, control failure, or exception expiration.

Publish exceptions in the conformance report; hidden exceptions invalidate the claim.

Critical exception cap

An unresolved exception against a critical control prevents Assured or Continuously Assured conformance and may prevent Operational conformance when the control score is below 3.

Source Authority and Freshness Register

NIST - Artificial Intelligence Risk Management Framework 1.0

Cross-sector AI risk governance, mapping, measurement, and management.

<https://doi.org/10.6028/NIST.AI.100-1>

NIST - NIST AI 600-1 - Generative Artificial Intelligence Profile

Generative-AI risks, controls, evaluation, and lifecycle actions.

<https://doi.org/10.6028/NIST.AI.600-1>

NIST - AI Resource Center and AI RMF Playbook

Current implementation resources and revision notices for the AI RMF.

<https://airc.nist.gov/>

ISO/IEC - ISO/IEC 42001:2023 - AI management systems

AI management-system requirements and continual improvement.

<https://www.iso.org/standard/42001>

ISO/IEC - ISO/IEC 23894:2023 - Guidance on AI risk management

AI risk-management guidance across lifecycle and stakeholders.

<https://www.iso.org/standard/77304.html>

ISO/IEC - ISO/IEC 42005:2025 - AI system impact assessment

Impact-assessment guidance for individuals and societies.

<https://www.iso.org/standard/42005>

MLCommons - MLCommons Benchmarks

Open performance, quality, and safety benchmark programs.

<https://mlcommons.org/benchmarks/>

Source Authority and Freshness Register - Continued

MLCommons - AI Risk and Reliability

Safety-test methodology and risk-oriented evaluation work.

<https://mlcommons.org/working-groups/ai-risk-reliability/ai-risk-reliability/>

Hugging Face - PEFT Documentation

Parameter-efficient fine-tuning methods and implementation guidance.

<https://huggingface.co/docs/peft/index>

Hugging Face - TRL Documentation

Post-training and preference-optimization framework documentation.

<https://huggingface.co/docs/trl/index>

Research - LoRA

Low-rank parameter-efficient adaptation.

<https://arxiv.org/abs/2106.09685>

Research - QLoRA

Quantized parameter-efficient fine-tuning.

<https://arxiv.org/abs/2305.14314>

Research - Direct Preference Optimization

Preference optimization without an explicit reward-model training loop.

<https://arxiv.org/abs/2305.18290>

NIST - SP 800-218A

Secure AI model-development lifecycle practices.

<https://csrc.nist.gov/pubs/sp/800/218/a/final>

Freshness rule

Sources were verified for this candidate edition on 2026-07-26. The standard **MUST** be revalidated when a cited authority changes, becomes superseded, is withdrawn, or materially alters the control interpretation.

Limitations, Revision History, and Adoption Position

This candidate Mythos standard is designed for engineering governance and assessment. It is not a legal opinion, regulatory certification, safety guarantee, or substitute for domain-specific professional judgment. Model behavior remains probabilistic and system risk depends on data, users, tools, deployment, incentives, and operating context.

Revision history

Version	Date	Change
2.0.0-candidate	2026-07-26	Expanded normative edition with 19 weighted controls, benchmark scenarios, conformance levels, machine-readable catalogs, assessment workbook integration, source freshness, and explicit lineage.
1.0.0-candidate	2026-07-24	Earlier permanent-identity candidate retained in the release lineage archive.

Adoption position

Use this edition as a candidate control and benchmark baseline. Organizations SHOULD pilot it on bounded systems, retain assessment evidence, submit corrections, and avoid representing candidate conformance as external certification.